Title

Complete genome sequence of the aquatic bacterium Runella slithyformis type strain (LSU 4(T)).

Journal

Standards in genomic sciences, 6(2)

ISSN

1944-3277

Authors

Copeland, Alex
Zhang, Xiaojing
Misra, Monica
et al.

Publication Date

2012-05-01

DOI

10.4056/sigs.2475579

Peer reviewed

Complete genome sequence of the aquatic bacterium *Runella slithyformis* type strain (LSU 4^T)

Alex Copeland[1], Xiaojing Zhang[1,2], Monica Misra[1,2], Alla Lapidus[1], Matt Nolan[1], Susan Lucas[1], Shweta Deshpande[1], Jan-Fang Cheng[1], Roxanne Tapia[1,2], Lynne A. Goodwin[1,2], Sam Pitluck[1], Konstantinos Liolios[1], Ioanna Pagani[1], Natalia Ivanova[1], Natalia Mikhailova[1], Amrita Pati[1], Amy Chen[3], Krishna Palaniappan[3], Miriam Land[1,4], Loren Hauser[1,4], Chongle Pan[1,4], Cynthia D. Jeffries[1,4], John C. Detter[1], Evelyne-Marie Brambilla[5], Manfred Rohde[6], Olivier D. Ngatchou Djao[6], Markus Göker[5], Johannes Sikorski[5], Brian J. Tindall[5], Tanja Woyke[1], James Bristow[1], Jonathan A. Eisen[1,7], Victor Markowitz[3], Philip Hugenholtz[1,8], Nikos C. Kyrpides[1], Hans-Peter Klenk[5]*, and Konstantinos Mavromatis[1]

[1] DOE Joint Genome Institute, Walnut Creek, California, USA
[2] Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA
[3] Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[4] Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
[5] Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany
[6] HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany
[7] University of California Davis Genome Center, Davis, California, USA
[8] Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia

*__Corresponding author__: Hans-Peter Klenk

Keywords: strictly aerobic, non-motile, Gram-negative, psychrotolerant, chemoorganotrophic, *Cytophagaceae*, *Cytophagia*, GEBA

Runella slithyformis Larkin and Williams 1978 is the type species of the genus *Runella*, which belongs to the *Cytophagaceae*, a family that was only recently classified to the order *Cytophagales* in the class *Cytophagia*. The species is of interest because it is able to grow at temperatures as low as 4°C. This is the first completed genome sequence of a member of the genus *Runella* and the sixth sequence from the family *Cytophagaceae*. The 6,919,729 bp long genome consists of a 6.6 Mbp circular genome and five circular plasmids of 38.8 to 107.0 kbp length, harboring a total of 5,974 protein-coding and 51 RNA genes and is a part of the ***G**enomic **E**ncyclopedia of **B**acteria and **A**rchaea* project.

Introduction

Strain LSU 4^T (= DSM 19594 = ATCC 29530 = NCIMB 11436) is the type strain of the species *Runella slithyformis*, which is the type species of its genus *Runella* [1,2]. The genus currently consists of four validly named species [3]. The genus name is derived from 'rune', a runic letter and the Latin diminutive ending 'ella', yielding the Neo-Latin word '*Runella*', meaning 'that which resembles figures of the runic alphabet' [3]. The species epithet is derived from slithy, a nonsense word from Lewis Carroll's Jabberwocky for a fictional organism that is 'slithy' and the Latin word 'suffix' meaning '-like, in the shape of', yielding the Neo-Latin word 'slithyformis' meaning 'slithy in form' [3]. *R. slithyformis* strain LSU 4^T was isolated from University Lake near Baton Rouge, Louisiana, USA, and described by Larkin and Williams in 1978 [1]. Another strain of *R. slithyformis*, termed strain 6, was isolated from Elbow Bayou near Baton Rouge [1]. Members of the genus *Runella* colonize diverse environmental habitats, preferentially aquatic ecosystems, including water bodies in Baton Rouge [1], a wastewater treatment plant in South-Korea [4], environmental water samples and their biofilms in

Japan [5], and an activated sludge process involved in enhanced biological removal of phosphor in Korea [6]. Another species of this genus was also isolated from the stems of surface-sterilized maize [7]. Here we present a summary classification and a set of features for *R. slithyformis* strain LSU 4^T, together with the description of the complete finished genome sequencing and annotation.

Classification and features

A representative genomic 16S rRNA sequence of *R. slithyformis* LSU 4^T was compared using NCBI BLAST [8,9] under default settings (e.g., considering only the high-scoring segment pairs (HSPs) from the best 250 hits) with the most recent release of the Greengenes database [10] and the relative frequencies of taxa and keywords (reduced to their stem [11]) were determined, weighted by BLAST scores. The most frequently occurring genera were *Runella* (31.0%), *Dyadobacter* (30.3%), *Cytophaga* (13.7%), *Cyclobacterium* (7.5%) and *Algoriphagus* (4.0%) (51 hits in total). Regarding the single hit to sequences from members of the species, the average identity within HSPs was 99.2%, whereas the average coverage by HSPs was 96.9%. Regarding the two hits to sequences from other members of the genus, the average identity within HSPs was 95.0%, whereas the average coverage by HSPs was 91.1%. Among all other species, the one yielding the highest score was *R. zeae* (NR_025004), which corresponded to an identity of 95.0% and an HSP coverage of 91.1%. (Note that the Greengenes database uses the INSDC (= EMBL/NCBI/DDBJ) annotation, which is not an authoritative source for nomenclature or classification.) The highest-scoring environmental sequence was GQ480089 ('changes during sewage treated process activated sludge wastewater treatment plant clone BXHB50'), which showed an identity of 96.6% and an HSP coverage of 98.0%. The most frequently occurring keywords within the labels of all environmental samples which yielded hits were 'skin' (5.5%), 'soil' (2.1%), 'sludg' (2.0%), 'biofilm' (1.7%) and 'forearm, volar' (1.7%) (199 hits in total). While few of these keywords fit the aquatic and sludge environments from which strain LSU 4^T originated, the majority of the hits point to human and even soil, which were, until now, not considered as habitats for *R. slithyformis*. However, environmental samples which yielded hits of a higher score than the highest scoring species were not found.

Figure 1 shows the phylogenetic neighborhood of *R. slithyformis* LSU 4^T in a 16S rRNA based tree. The sequences of the two identical 16S rRNA gene copies in the genome do not differ from the previously published 16S rRNA sequence (M62786), which contains 13 ambiguous base calls.

The cells of strain LSU 4^T are generally curved rods, with the degree of curvature of individual cells within a culture varying from nearly straight to crescent shape. Cell diameter varies from 0.5 to 0.9 μm, and the length from 2.0 to 3.0 μm [1]. With its curved rod shape, strain LSU 4^T differs from other members of the genus, such as *R. limosa* which has long rods while *R. zeae* is bent rod-shaped [6]. On the MS agar medium used at the time of isolation, *R. slithyformis* rarely formed long spirals. However, Chelius and Triplett [23] reported the formation of long spirals by the strain LSU 4^T when cells were allowed to grow in R2A broth medium (see Figure 2). Larkin and Williams [1] reported a possible production of filaments up to 14 μm in length, which are not coiled. This contrasts the findings of Chelius *et al*. [7] who described the cells of the strain LSU 4^T as circular with swollen ends that would not form filaments. Rings with an outer diameter of 2.0 to 3.0 μm may also occur [1]. Colonies produced a pale pink, nondiffusible, nonfluorescent pigment on MS agar [1]. The strain LSU 4^T is a Gram-negative bacterium (Table 1). Strain LSU 4^T is non-motile, aerobic and chemoorganotrophic [1]. It does not grow on media with NaCl concentrations of 1.5% or higher [23]. This feature was similar to that of another member of this genus, *R. zeae* [7]. The temperature range for growth is between 4°C-37°C, with an optimum between 20°C-30°C [6]; the strain being unable to grow at temperatures above 37°C [23]. The sole carbon sources used by the strain LSU 4^T for growth on MS agar are glycogen, D-arabitol, dulcitol, inositol, mannitol, sorbitol and sorbose, but the growth was weak except in the presence of glycogen [23]. Some of these features are however contradictory to the findings of Chelius *et al*. [7] whose attempt to grow the strain LSU 4^T in the presence of glycogen in R2A medium was unsuccessful. Further detailed physiological insight, e.g., carbon source utilization in R2A medium, MS agar medium, or by the API 50 CH test, have been reported previously [7,23]. Also, resistance to a variety of antibiotics has been reported [7,23].

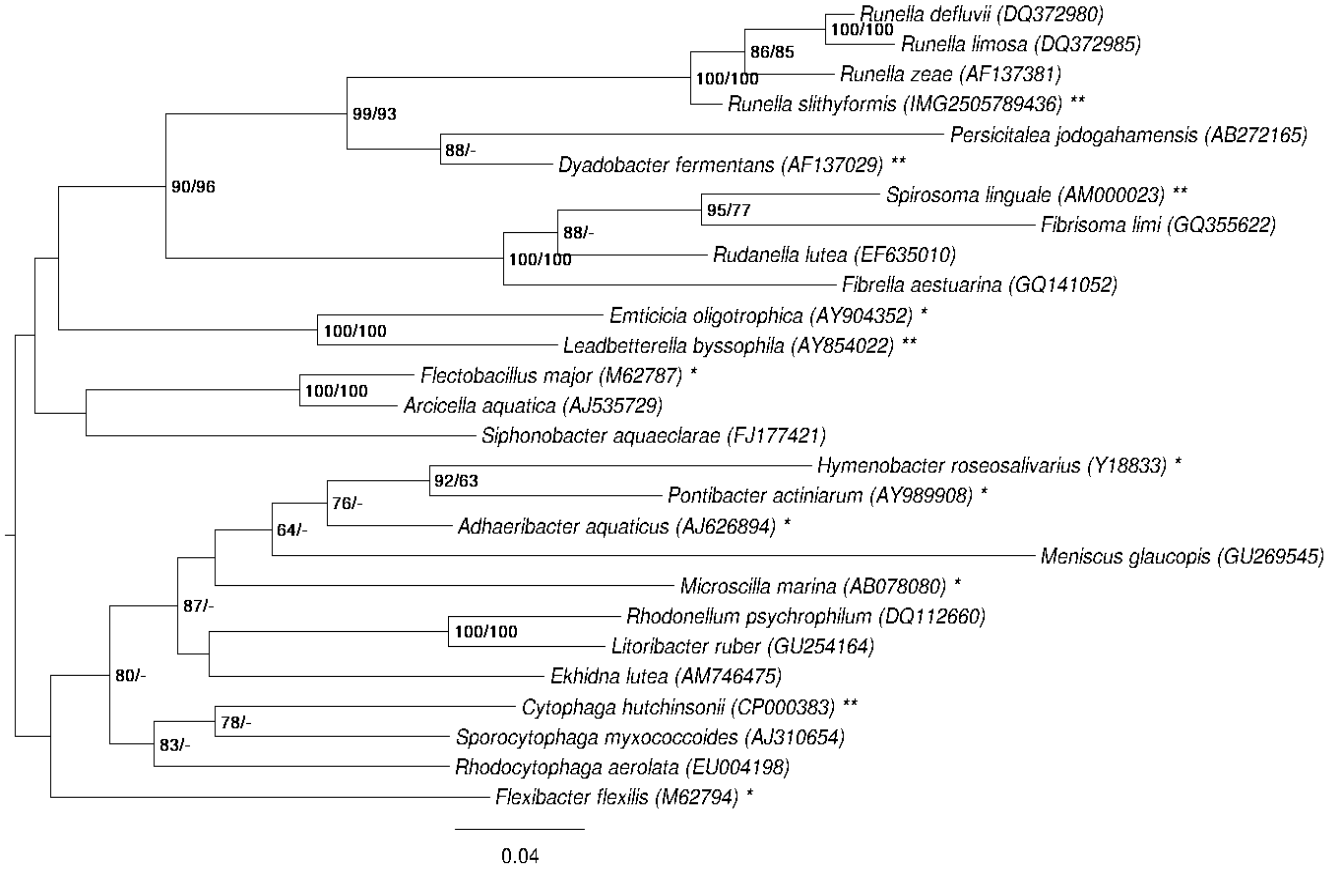

Figure 1. Phylogenetic tree highlighting the position of *R. slithyformis* relative to the type strains of the type species of the other genera within the family *Cytophagaceae*. The tree was inferred from 1,330 aligned characters [12,13] of the 16S rRNA gene sequence under the maximum likelihood (ML) criterion [14]. Rooting was done initially using the midpoint method [15] and then checked for its agreement with the current classification (Table 1). The branches are scaled in terms of the expected number of substitutions per site. Numbers adjacent to the branches are support values from 400 ML bootstrap replicates [16] (left) and from 1,000 maximum parsimony bootstrap replicates [17] (right) if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [18] are labeled with one asterisk, those also listed as 'Complete and Published' with two asterisks [19-22].

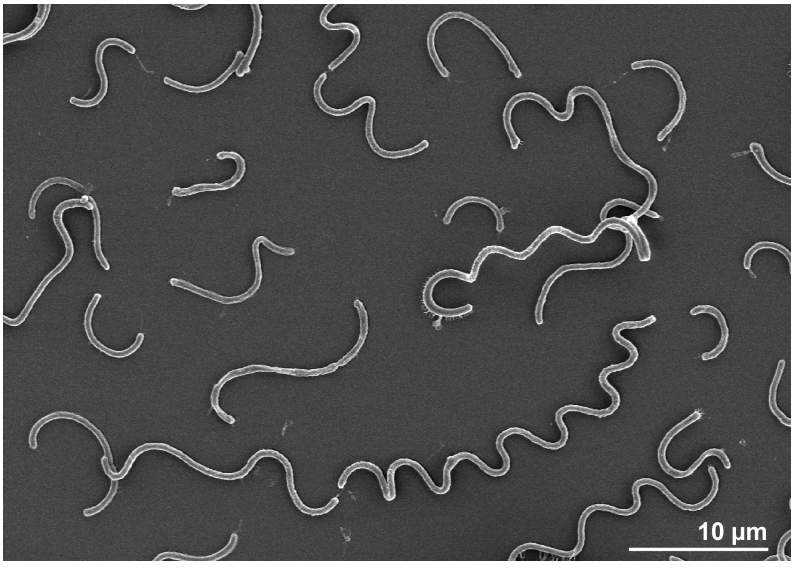

Figure 2. Scanning electron micrograph of *R. slithyformis* LSU 4^T

Chemotaxonomy

The principal cellular fatty acids of strain LSU 4^T are *iso*-$C_{15:0\ 2\text{-}OH}$/$C_{16:1}\omega7c$ (32.1%), *iso*-$C_{15:0}$ (19.8%) and $C_{16:1}\omega5c$ (16.5%) [23]. Minor fatty acids include $C_{16:0}$ (7.1%), *iso*-$C_{17:0\ 3\text{-}OH}$ (7.0%), *anteiso*-$C_{15:0}$ (4.3%), *iso*-$C_{15:0\ 3\text{-}OH}$ (4.1%), *iso*-$C_{15:1}$ G (2.4%), $C_{16:0\ 3\text{-}OH}$ (2.0%), an unknown one (ECL 13.6) (1.83%) and $C_{15:0}$ (1.5%) [23]. Major polar lipids were not reported for strain LSU 4^T, but those of the genus *Runella* could be retrieved from *R. defluvii* strain EMB13^T and *R. limosa* strain EMB111^T [4,6].

Table 1. Classification and general features of *R. slithyformis* LSU 4^T according to the MIGS recommendations [24].

MIGS ID	Property	Term	Evidence code
	Current classification	Domain *Bacteria*	TAS [25]
		Phylum *Bacteroidetes*	TAS [26,27]
		Class *Cytophagia*	TAS [27,28]
		Order *Cytophagales*	TAS [2,29]
		Family *Cytophagaceae*	TAS [2,30]
		Genus *Runella*	TAS [1,2]
		Species *Runella slithyformis*	TAS [1,2]
MIGS-7		Type strain LSU 4^T	TAS [1,2]
	Gram stain	negative	TAS [1]
	Cell shape	curved rod-shaped, rigid	TAS [1]
	Motility	non-motile	TAS [1]
	Sporulation	none	TAS [1]
	Temperature range	psychrotolerant mesophiles, grows at temperatures as low as 4°C	TAS [23]
	Optimum temperature	20°C-30°C	TAS [6]
	Salinity	no growth in the presence of NaCl (1.5%)	TAS [31]
MIGS-22	Relationship to oxygen	strictly aerobic	TAS [1]
	Carbon source	carbohydrates	TAS [1,23]
	Energy metabolism	chemoorganotroph	TAS [1]
MIGS-6	Habitat	fresh water	TAS [1]
MIGS-15	Biotic relationship	free living	NAS
MIGS-14	Known pathogenicity	none	NAS
MIGS-16	Specific host	none	NAS
	Biosafety level	1	TAS [32]
MIGS-23.1	Isolation	fresh water lake	TAS [1]
MIGS-4	Geographic location	University Lake, Baton Rouge, Louisiana, USA	TAS [1]
MIGS-5	Time of sample collection	1978 or before	TAS [1]
MIGS-4.1	Latitude	30.417	NAS
MIGS-4.2	Longitude	-91.167	NAS
MIGS-4.3	Depth	not reported	
MIGS-4.4	Altitude	15 m	NAS

Evidence codes - IDA: Inferred from Direct Assay (first time in publication); TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project. If the evidence code is IDA, then the property was directly observed for a living isolate by one of the authors or an expert mentioned in the acknowledgements [33].

Genome sequencing and annotation

Genome project history

This organism was selected for sequencing on the basis of its phylogenetic position [34], and is part of the ***G**enomic **E**ncyclopedia of **B**acteria and **A**rchaea* project [35]. The genome project is deposited in the Genomes On Line Database [18] and the complete genome sequence is deposited in GenBank. Sequencing, finishing and annotation were performed by the DOE Joint Genome Institute (JGI). A summary of the project information is shown in Table 2.

Table 2. Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	Four genomic libraries: one 454 pyrosequence standard library, two 454 PE libraries (2 kb and 11 kb insert sizes), one Illumina library
MIGS-29	Sequencing platforms	Illumina GAii, 454 GS FLX Titanium
MIGS-31.2	Sequencing coverage	100.4 × Illumina; 28.2 × pyrosequence
MIGS-30	Assemblers	Newbler version 2.3, Velvet 0.7.63, phrap version SPS - 4.24
MIGS-32	Gene calling method	Prodigal 1.4, GenePRIMP
	INSDC ID	CP002859 (chromosome) CP002860-64 (plasmids RUNSL01-05)
	Genbank Date of Release	August 16, 2011
	GOLD ID	Gc01829
	NCBI project ID	49125
	Database: IMG-GEBA	2505679030
MIGS-13	Source material identifier	DSM 19594
	Project relevance	Tree of Life, GEBA

Growth conditions and DNA isolation

R. slithyformis strain LSU 4^T, DSM 19594, was grown in DSMZ medium 7 (*Ancyclobacter-Spirosoma* medium) [36] at 28°C. DNA was isolated from 0.5-1 g of cell paste using MasterPure Gram-positive DNA purification kit (Epicentre MGP04100) following the standard protocol as recommended by the manufacturer with modification st/DL for cell lysis as described in Wu *et al.* 2009 [35]. DNA is available through the DNA Bank Network [31].

Genome sequencing and assembly

The genome was sequenced using a combination of Illumina and 454 sequencing platforms. All general aspects of library construction and sequencing can be found at the JGI website [37]. Pyrosequencing reads were assembled using the Newbler assembler (Roche). The initial Newbler assembly consisting of 121 contigs in two scaffolds was converted into a phrap [38] assembly by making fake reads from the consensus, to collect the read pairs in the 454 paired end library. Illumina GAii sequencing data (638.9 Mb) was assembled with Velvet [39] and the consensus sequences were shredded into 2.0 kb overlapped fake reads and assembled together with the 454 data. The 454 draft assembly was based on 206.2 Mb 454 draft data and all of the 454 paired end data. Newbler parameters are -consed -a 50 -l 350 -g -m -ml 20. The Phred/Phrap/Consed software package [38] was used for sequence assembly and quality assessment in the subsequent finishing process. After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with gapResolution [37], Dupfinisher [40], or sequencing cloned bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks (J.-F. Chang, unpublished). A total of 289 additional reactions and 3 shatter libraries

were necessary to close gaps and to raise the quality of the finished sequence. Illumina reads were also used to correct potential base errors and increase consensus quality using a software Polisher developed at JGI [41]. The error rate of the completed genome sequence is less than 1 in 100,000. Together, the combination of the Illumina and 454 sequencing platforms provided 128.6 × coverage of the genome. The final assembly contained 540,807 pyrosequence and 19,068,176 Illumina reads.

Genome annotation

Genes were identified using Prodigal [42] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [43]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGR-Fam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation was performed within the Integrated Microbial Genomes - Expert Review (IMG-ER) platform [44].

Genome properties

The genome consists of one circular chromosome with a length of 6,568,739 bp and a G+C content of 47%, and five circular plasmids with 38,784 bp, 44,754 bp, 66,926 bp, 93,527 bp and 106,999 bp length, respectively (Table 3 and Figure 3). Of the 6,025 genes predicted, 5,974 were protein-coding genes, and 51 RNAs; 182 pseudogenes were also identified. The majority of the protein-coding genes (59.7%) were assigned a putative function while the remaining ones were annotated as hypothetical proteins. The distribution of genes into COGs functional categories is presented in Table 4.

Table 3. Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	6,919,729	100.00%
DNA coding region (bp)	6,063,039	87.62%
DNA G+C content (bp)	3,212,364	46.42%
Number of replicons	6	
Extrachromosomal elements	5	
Total genes	6,025	100.00%
RNA genes	51	0.85%
rRNA operons	2	
tRNA genes	43	0.71%
Protein-coding genes	5,974	99.15%
Pseudo genes	182	3.02%
Genes with function prediction	3,599	59.73%
Genes in paralog clusters	3,238	53.74%
Genes assigned to COGs	3,912	64.93%
Genes assigned Pfam domains	4,008	66.52%
Genes with signal peptides	1,748	29.01%
Genes with transmembrane helices	1,350	22.41%
CRISPR repeats	0	

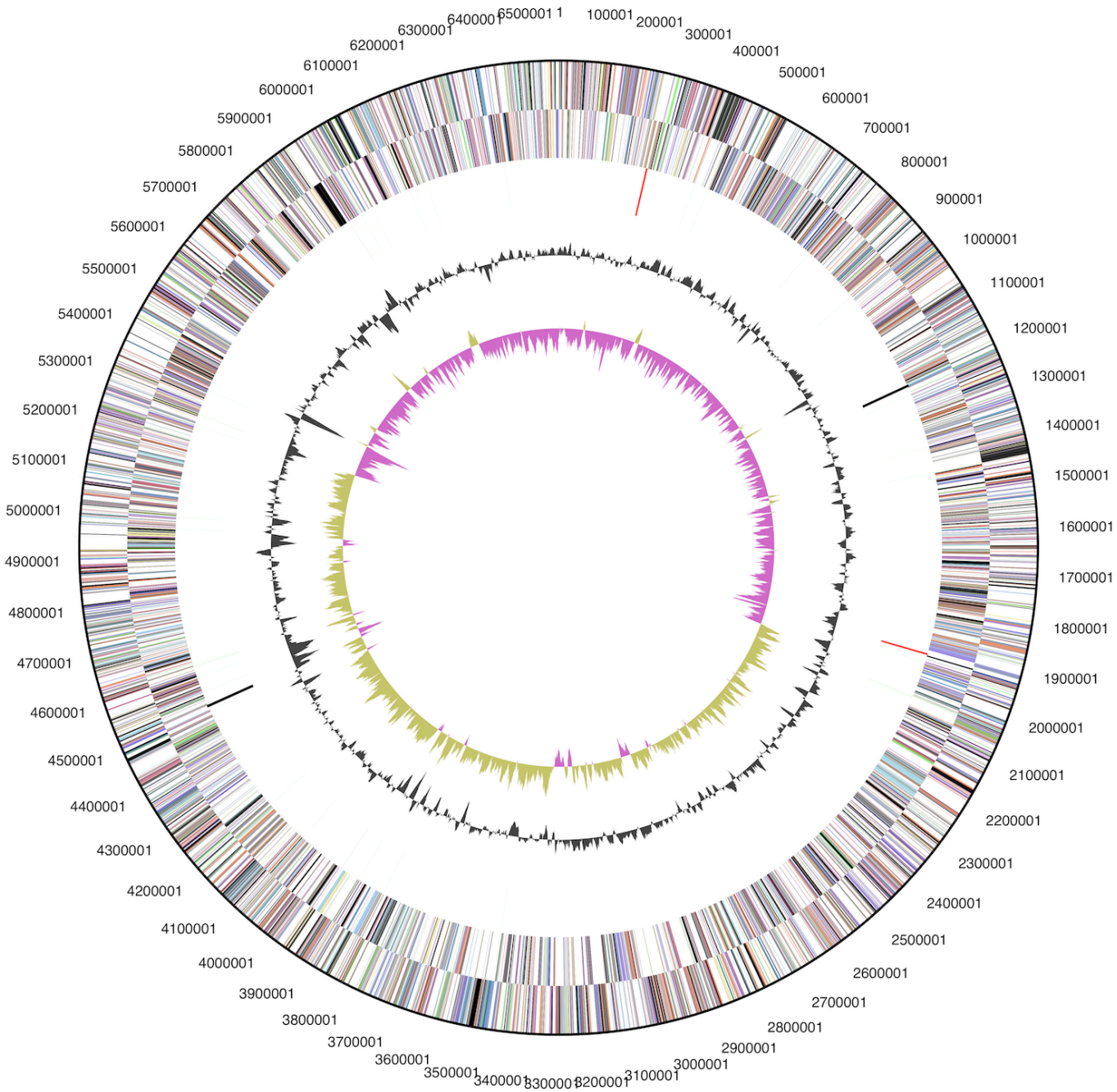

Figure 3. Graphical map of the circular chromosome (plasmids not shown, but accessible through the img/er pages on the JGI web pages [37]). From outside to center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew.

Table 4. Number of genes associated with the general COG functional categories

Code	value	%age	Description
J	173	4.0	Translation, ribosomal structure and biogenesis
A	0	0.0	RNA processing and modification
K	338	7.8	Transcription
L	216	5.0	Replication, recombination and repair
B	1	0.2	Chromatin structure and dynamics
D	36	0.8	Cell cycle control, cell division, chromosome partitioning
Y	0	0.0	Nuclear structure
V	126	2.9	Defense mechanisms
T	272	6.3	Signal transduction mechanisms
M	372	8.6	Cell wall/membrane/envelope biogenesis
N	14	0.3	Cell motility
Z	1	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	73	1.7	Intracellular trafficking, secretion, and vesicular transport
O	132	3.1	Posttranslational modification, protein turnover, chaperones
C	204	4.7	Energy production and conversion
G	351	8.1	Carbohydrate transport and metabolism
E	304	7.0	Amino acid transport and metabolism
F	86	2.0	Nucleotide transport and metabolism
H	161	3.7	Coenzyme transport and metabolism
I	158	3.7	Lipid transport and metabolism
P	226	5.2	Inorganic ion transport and metabolism
Q	99	2.3	Secondary metabolites biosynthesis, transport and catabolism
R	620	14.3	General function prediction only
S	372	8.6	Function unknown
-	2,113	35.1	Not in COGs

Acknowledgements

We would like to gratefully acknowledge the help of Regine Fähnrich (DSMZ) for growing *R. slithyformis* cultures. This work was performed under the auspices of the US Department of Energy Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396, UT-Battelle and Oak Ridge National Laboratory under contract DE-AC05-00OR22725, as well as German Research Foundation (DFG) INST 599/1-2.

References

1. Larkin JM, Williams PM. *Runella slithyformis* gen. nov., sp. nov., a curved, nonflexible, pink bacterium. *Int J Syst Bacteriol* 1978; **28**:32-36. http://dx.doi.org/10.1099/00207713-28-1-32

2. Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. *Int J Syst Bacteriol* 1980; **30**:225-420. http://dx.doi.org/10.1099/00207713-30-1-225

3. Euzéby JP. List of bacterial names with standing in nomenclature: a folder available on the Internet. *Int J Syst Bacteriol* 1997; **47**:590-592. PubMed http://dx.doi.org/10.1099/00207713-47-2-590

4. Lu S, Lee JR, Ryu SH, Chung BS, Choe WS, Jeon CO. *Runella defluvii* sp. nov., isolated from a domestic wastewater treatment plant. *Int J Syst Evol Microbiol* 2007; **57**:2600-2603. PubMed http://dx.doi.org/10.1099/ijs.0.65252-0

5. Furuhata K, Kato Y, Goto K, Saitou K, Sugiyama J, Hara M, Fukuyama M. Identification of pink-pigmented bacteria isolated from environmental water samples and their biofilm fromation abilities. *Biocontrol Sci* 2008; **13**:33-39. PubMed http://dx.doi.org/10.4265/bio.13.33

6. Ryu SH, Nguyen TTH, Park W, Kim CJ, Jeon CO. *Runella limosa* sp. nov., isolated from activated sludge. *Int J Syst Evol Microbiol* 2006; **56**:2757-2760. PubMed http://dx.doi.org/10.1099/ijs.0.64460-0

7. Chelius MK, Henn JA, Triplett EW. *Runella zeae* sp. nov., a novel Gram-negative bacterium from the stems of surface-sterilized *Zea mays*. *Int J Syst Evol Microbiol* 2002; **52**:2061-2063. PubMed http://dx.doi.org/10.1099/ijs.0.02203-0

8. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**:403-410. PubMed

9. Korf I, Yandell M, Bedell J. BLAST, O'Reilly, Sebastopol, 2003.

10. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006; **72**:5069-5072. PubMed http://dx.doi.org/10.1128/AEM.03006-05

11. Porter MF. An algorithm for suffix stripping. *Program: electronic library and information systems*. 1980; **14**:130-137.

12. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; **17**:540-552. PubMed http://dx.doi.org/10.1093/oxfordjournals.molbev.a026334

13. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002; **18**:452-464. PubMed http://dx.doi.org/10.1093/bioinformatics/18.3.452

14. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web-servers. *Syst Biol* 2008; **57**:758-771. PubMed http://dx.doi.org/10.1080/10635150802429642

15. Hess PN, De Moraes Russo CA. An empirical test of the midpoint rooting method. *Biol J Linn Soc Lond* 2007; **92**:669-674. http://dx.doi.org/10.1111/j.1095-8312.2007.00864.x

16. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? *Lect Notes Comput Sci* 2009; **5541**:184-200. http://dx.doi.org/10.1007/978-3-642-02008-7_13

17. Swofford DL. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0 b10. Sinauer Associates, Sunderland, 2002.

18. Liolios K, Chen IM, Mavromatis K, Tavernarakis N, Hugenholtz P, Markowitz VM, Kyrpides NC. The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2010; **38**:D346-D354. PubMed http://dx.doi.org/10.1093/nar/gkp848

19. Abt B, Teshima H, Lucas S, Lapidus A, Del Rio TG, Nolan M, Tice H, Cheng JF, Pitluck S, Liolios K, *et al*. Complete genome sequence of *Leadbetterella byssophila* type strain (4M15^T). *Stand Genomic Sci* 2011; **4**:2-12. PubMed http://dx.doi.org/10.4056/sigs.1413518

20. Lail K, Sikorski J, Saunders E, Lapidus A, Del Rio TG, Copeland A, Tice H, Cheng JF, Lucas S, Nolan M, *et al*. Complete genome sequence of *Spirosoma linguale* type strain (1^T). *Stand Genomic Sci* 2010; **2**:176-185. PubMed http://dx.doi.org/10.4056/sigs.741334

21. Lang E, Lapidus A, Chertkov O, Brettin T, Detter JC, Han C, Copeland A, Del Rio TG, Nolan M, Chen F, *et al*. Complete genome sequence of *Dyadobacter fermentans* type strain (NS114^T). *Stand Genomic Sci* 2009; **1**:133-140. PubMed http://dx.doi.org/10.4056/sigs.19262

22. Xie G, Bruce DC, Challacombe JF, Chertkov O, Detter JC, Gilna P, Han CS, Lucas S, Misra M, Myers GL, *et al*. Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*. *Appl Environ Microbiol* 2007; **73**:3536-3546. PubMed http://dx.doi.org/10.1128/AEM.00225-07

23. Chelius MK, Triplett EW. gen. nov., sp. nov., a novel Gram-negative bacterium isolated from surface-sterilized *Zea mays* stems. *Int J Syst Evol Microbiol* 2000; **50**:751-758. PubMed http://dx.doi.org/10.1099/00207713-50-2-751

24. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, *et al*. The minimum information about a genome sequence (MIGS) specification.

Nat Biotechnol 2008; **26**:541-547. PubMed http://dx.doi.org/10.1038/nbt1360

25. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains, and *Eucarya. Proc Natl Acad Sci USA* 1990; **87**:4576-4579. PubMed http://dx.doi.org/10.1073/pnas.87.12.4576

26. Krieg NR, Ludwig W, Euzéby J, Whitman WB. Phylum XIV. *Bacteroidetes* phyl. nov. *In:* Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, Ludwig W, Whitman WB (*eds*), Bergey's Manual of Systematic Bacteriology, second edition, vol. 4 (The *Bacteroidetes*, Spirochaetes, Tenericutes (*Mollicutes*), *Acidobacteria*, Fibrobacteres, *Fusobacteria*, Dictyoglomi, *Gemmatimonadetes*, *Lentisphaerae*, Verrucomicrobia, *Chlamydiae*, and Planctomycetes), Springer, New York, 2010, p. 25.

27. List Editor. Validation List N° 143. *Int J Syst Evol Microbiol* 2012; **62**:1-4. http://dx.doi.org/10.1099/ijs.0.039487-0

28. Nakagawa Y. Class IV. *Cytophagia* class. nov. *In*: Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, Ludwig W, Whitman WB (*eds*), Bergey's Manual of Systematic Bacteriology, second edition, vol. 4 (The *Bacteroidetes*, *Spirochaetes*, *Tenericutes* (*Mollicutes*), *Acidobacteria*, *Fibrobacteres*, *Fusobacteria*, *Dictyoglomi*, *Gemmatimonadetes*, *Lentisphaerae*, *Verrucomicrobia*, *Chlamydiae*, and *Planctomycetes*), Springer, New York, 2010, p. 370.

29. Leadbetter ER. Order II. *Cytophagales* Nomen novum. *In*: Buchanan RE, Gibbons NE (*eds*), Bergey's Manual of Determinative Bacteriology, Eighth Edition, The Williams and Wilkins Co., Baltimore, 1974, p. 99.

30. Stanier RY. Studies on the Cytophagas. *J Bacteriol* 1940; **40**:619-635. PubMed

31. Gemeinholzer B, Dröge G, Zetzsche H, Haszprunar G, Klenk HP, Güntsch A, Berendsohn WG, Wägele JW. The DNA Bank Network: the start from a German initiative. *Biopreserv Biobank* 2011; **9**:51-55. http://dx.doi.org/10.1089/bio.2010.0029

32. BAuA. 2010. Classification of bacteria and archaea in risk groups. TRBA 466, p. 194. www.baua.de Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Germany.

33. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000; **25**:25-29. PubMed http://dx.doi.org/10.1038/75556

34. Klenk HP, Göker M. En route to a genome-based classification of and*? Syst Appl Microbiol* 2010; **33**:175-182. PubMed http://dx.doi.org/10.1016/j.syapm.2010.03.003

35. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, *et al*. A phylogeny-driven genomic encyclopaedia of and. *Nature* 2009; **462**:1056-1060. PubMed http://dx.doi.org/10.1038/nature08656

36. List of growth media used at DSMZ: http://www.dsmz.de/catalogues/catalogue-microorganisms/culture-technology/list-of-media-for-microorganisms.html.

37. JGI website. http://www.jgi.doe.gov

38. The Phred/Phrap/Consed software package. http://www.phrap.com

39. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; **18**:821-829. PubMed http://dx.doi.org/10.1101/gr.074492.107

40. Han C, Chain P. Finishing repeat regions automatically with Dupfinisher. *In:* Proceeding of the 2006 international conference on bioinformatics & computational biology. Arabnia HR, Valafar H (*eds*), CSREA Press. June 26-29, 2006: 141-146.

41. Lapidus A, LaButti K, Foster B, Lowry S, Trong S, Goltsman E. POLISHER: An effective tool for using ultra short reads in microbial genome assembly and finishing. AGBT, Marco Island, FL, 200845. *J Bacteriol* 1940; **40**:619-635.

42. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**:119. PubMed http://dx.doi.org/10.1186/1471-2105-11-119

43. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 2010; **7**:455-457. PubMed http://dx.doi.org/10.1038/nmeth.1457

44. Markowitz VM, Ivanova NN, Chen IMA, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278. PubMed http://dx.doi.org/10.1093/bioinformatics/btp393